

Afnan Amin Ali

• +92-34-23428364 • afnanaminali@gmail.com • [Karachi, Pakistan](#) • [LinkedIn](#) • [Github](#)

WORK EXPERIENCE

Bazaar

Nov 2021-Present

Data Engineer

Karachi, Pakistan

- Designed and implemented various production ETL pipelines using Apache Airflow (for job scheduling and orchestration) and AWS EMR (for compute) which are used for reporting and client facing analytical purposes.
- Built a scalable and reliable Data Warehouse to extract the data using AWS DMS, transform and load into AWS S3 buckets in ORC format and developed data products consuming this data which helped drive crucial business and product decisions.
- Integrated Apache Hudi with a data platform for performing ACID properties on our terabytes of data.
- Designed and implemented a real-time data pipeline to process semi-structured data by integrating 150 million raw records from different data sources using Pyspark.
- Ingested data from disparate data sources using a combination of SQL, Google Analytics API, and Salesforce API using Python to create data views to be used in BI tools like Tableau.
- Designed SheetBots and Slackbots in PySpark for data updates every hour and automated them using Airflow.
- Researched and integrated various data tools including Airbyte, Open Metadata for row level security in our data platform.

TimeXperts

Jun 2021-Nov 2021

Associate Software Engineer

Karachi, Pakistan

- Migrated from older ERP versions to newer versions for various clients including National Bank of Canada and Toyota which enabled them to use new features and bug fixes.

TPS

Oct 2020-Jun 2021

Associate Software Engineer

Karachi, Pakistan

- Developed critical transactions flows in C++ following different mandates from various payment providers including Mastercard, VISA and CUP for different banks.
- Optimized Oracle MySQL stored procedures and queries to reduce CPU consumption.
- Investigated and worked on Ad-hoc banking issues, such as lending, liability, zakat deduction and customer creation.
- Wrote BASH scripts to upload images on ATM following custom datamaps provided by vendors.

TECHNOLOGIES AND LANGUAGES

Languages: Python, Pyspark, C++, Javascript, PHP

Technologies: Docker, Trino, Linux, AWS (S3, EC2, EKS, DMS, DynamoDB, Glue), Apache Hive, Apache Superset, Apache Hudi, Apache Airflow, Grafana, OpenMetaData, Airbyte, MySQL, MongoDB, Postgres

EDUCATION

NED University Of Engineering and Technology

2016–2020

BE, Computer System Engineering

3.72/4.0 CGPA

- Conducted webinar on Computer Vision using Tensorflow under the umbrella of [AI Club](#), and handled public relations to invite various industry professionals to conduct various workshops.

PROJECTS

- Developed a big data application for querying and visualization of data and integrated Trino as a distributed engine for handling queries which consumes large amounts of memory.
- Designed a banking application that authenticates users based on their fingerprints.
- Implemented a custom protocol to optimize delay in BRAM for efficient FPGA image processing.
- Developed website that grows plants in university map using Mapbox Api